Digital Approaches to Analyzing and Translating Emotion : What Is Love?

Alstola, Tero

Routledge
2022-08-30

Alstola , T , Jauhiainen , H , Svärd , S , Sahala , A & Linden , K 2022 , Digital Approaches to Analyzing and Translating Emotion : What Is Love? in K Sonik & U Steinert (eds) , The Routledge Handbook of Emotions in the Ancient Near East . Routledge , London , pp.
þÿ 8 8 1 1 6 . h t t p s : / / d o i . o r g / 1 0 . 4 3 2 4 / 9 7 8 0 3 6 7 8 2 2 8 7 3 - 6

http://hdl.handle.net/10138/348398
https://doi.org/10.4324/9780367822873-6

3

DIGITAL APPROACHES TO ANALYZING AND TRANSLATING EMOTION

What Is Love?

Tero Alstola, Heidi Jauhiainen, Saana Svärd, Aleksi Sahala, and Krister Lindén

Introduction

The field of language technology is booming, thanks to rapidly growing digital data and the development of new methods and tools to analyze it.[1] Anyone using speech recognition software, online translation services, or a grammar checker is familiar with concrete applications of language-technological research. Methods and tools from language technology can also be applied to ancient texts when digitized text corpora are large enough. In this study, we focus on the analysis of emotion in Akkadian texts using *râmu*, "to love," as a case study.

This chapter applies two language-technological methods, pointwise mutual information (PMI) and the fastText implementation of the continuous skip-gram model, to a dataset of 7,346 Akkadian cuneiform texts from the Open Richly Annotated Cuneiform Corpus (Oracc). These texts were written primarily in the Neo-Assyrian Period (934–612 BCE) in Assyria and Babylonia, but earlier and later texts are also included. The texts belong to several genres, ranging from letters and royal inscriptions to legal and administrative texts.

PMI and the continuous skip-gram model can be used to study the semantic domains in which lexemes – in this case, emotion words – occur. PMI detects words which typically co-occur in the dataset: for example, "to fear" may co-occur with "dark," "spider," and "panic." The continuous skip-gram model finds words which appear in similar semantic contexts: for example, "to be angry," "to rage," and "to be furious" are words which are not necessarily used together but are likely to appear in similar contexts. To illustrate the potential of these methods, we apply them to analyze the semantic domains of the verb *râmu*, "to love," and its derivatives in Akkadian. The usage and semantic domains of a word can vary greatly between different genres. As our dataset consists of several genres, we focus the analysis on royal inscriptions, letters, and literary text genres. All our research data is openly available online at https://doi.org/10.5281/zenodo.5861579.

We begin by giving an overview of current digital research on emotions, and we outline the availability of ancient textual data that can be used for a language-technological study of emotions. This is followed by a discussion of our dataset and methods. Finally, our case study highlights how the usage of *râmu* and its derivatives varies across different genres but can at

 DOI: 10.4324/9780367822873-6

the same time be rather stable within a certain genre. Like the word *love* in English, *râmu* can denote different aspects of affection and love. It refers, for example, to erotic and sexual relationships between people, affection between family members, the king's love of justice, and the gods' pleasure with and acceptance of the king who fulfills divine expectations.

Digital Scholarship of Emotions

Methods and Approaches

There is in existence a broad array of digital approaches to emotions, and the methods used in this chapter cover only a certain niche in the field. In computational linguistics, sentiment analysis has been very popular in recent years. In sentiment analysis, a text is analyzed to detect positive or negative opinions and emotions (e.g., Ge et al. 2018). It is used, for example, to study emotions in social media and by marketers to study consumer opinions. Sentiment analysis does not always have to do with emotions per se, whereas "emotion recognition" does (Schnoebelen 2012, 23–26). The detection of emotions from vocal expressions was first explored by psychologists; computational methods have since been used to study all kinds of acoustic measures in order to recognize emotions in speech. From text, emotions can be detected, for example, by using lexicons or by studying linguistic features by means of machine learning (Canales and Martínez-Barco 2014).

We are interested in the semantic domains of emotion words and how they can be analyzed with computational methods and visualized as linguistic networks. In addition to our methods and workflow, there are many other fruitful approaches to these very same questions. Toivonen et al. (2012) studied the similarity of various emotion words in Finnish. They used human-annotated similarities between 50 words and built networks which were visualized in several ways. By identifying clusters of adjacent triangles and specific local network structures, they could shed light on why human annotators consider certain emotions similar. Jackson et al. (2019) studied the "colexification" of emotions, that is, how the same word is associated with two or more emotion concepts. They built a database of colexifications for 2,474 languages and formed colexification networks within 20 language families, connecting emotion words by colexifications. They found that greater variation in semantics can partly be explained by greater geographical distance between language families and that there is "evidence for a common underlying structure in the meaning of emotion concepts across languages" (Jackson et al. 2019, 1521). The valence and physiological activation associated with emotions predict the structure of colexification networks across different language families.

Availability of and Prerequisites for Data

Language-technological analysis of emotion words requires large digitized text corpora, which are increasingly available for ancient languages as well as for modern ones. Certain computational methods can work well with a dataset of several hundred texts or tens of thousands of words, but they generally perform better with corpora of thousands of texts or millions of words. The annotation of a dataset also affects its suitability for language-technological analysis. In highly inflected languages such as Akkadian or Greek, transliterated or transcribed text is not well suited for digital analysis, as a single word can appear in numerous forms in the text. A text annotated with lemmas (dictionary forms) of the words is needed. For Akkadian and Sumerian, the Open Richly Annotated Cuneiform Corpus provides thousands of texts with rich metadata, including transliteration, transcription, lemmas, translation, and part-of-speech

tagging.[2] Similar corpora are available for other ancient languages, including Greek, Latin, and biblical Hebrew.[3] However, these corpora are of unequal size: as of December 2020, Oracc contains more than 2 million words in Akkadian and Sumerian and the Scaife Viewer 30 million words in Greek and 17 million in Latin, but the word count of the Hebrew Bible is only circa 300,000. Accordingly, methods that yield good results for the Greek and Latin corpora may not be applicable to biblical Hebrew because of the difference in word counts.

Dataset for This Chapter

Our data was downloaded from Oracc in the form of JSON files in February 2019. For the analysis of the word *râmu*, "to love," we used a dataset consisting of 7,346 Akkadian texts. All texts tagged as written in "Akkadian," "Neo-Assyrian," or "Neo-Babylonian" are included. We also chose to use texts in which all of the words were marked as being in Akkadian, even if the language of the overall text was not indicated in the Oracc metadata. Bilingual (e.g., Sumerian-Akkadian) texts are not included, though the texts utilized may contain several different dialects of Akkadian.[4] The dataset represents many different text genres, the largest group of which is formed by different kinds of letters (2,247 texts). Royal inscriptions (1,494) and legal transactions (1,404) are two other prominent genres among the texts. More than 76% of the texts are from the Neo-Assyrian Period (5,638), and over half of them (63%) originate from the city of Nineveh (4,646).

We pre-processed the texts to adapt them for computational analysis. We standardized the spellings of divine and place names and removed duplicate texts following the procedure explained in Alstola et al. (2019, 162–63). We only used lemmas (dictionary forms) of content words – nouns, verbs, and adjectives – while we replaced all the other words with an underscore as a placeholder. Since neither the cuneiform script nor the Oracc metadata indicates sentence endings, each document is treated as one continuous line of text. The transcription of Akkadian lemmas and their translations in this article primarily follow the *Concise Dictionary of Akkadian* (CDA), which is also the recommended source of lemmas and their translations in Oracc. The translations of lemmas in our online dataset may differ from the translations used in this chapter, because different Oracc projects are not fully consistent in this regard.

In our previous work (Svärd et al. 2021a), we combined all the derivatives of an emotion verb under one word to facilitate its computational analysis. The more often a word appears in the dataset, the more contextual information our computational methods have to analyze its collocates and semantic contexts. We noticed, however, that this approach often obscures important differences in how the derivatives of a given verb are used. It has also become clear that genre plays a decisive role in the usage and semantics of a word. Consequently, we treat the emotion verb and its derivatives as separate words in this chapter and analyze their attestations in different genres. We use the same dataset we used in Svärd et al. (2021a; see also Svärd et al. 2021b) but, as outlined in the following, we did some additional processing on the specific emotion words that we are interested in. We focus on the verb *râmu*, "to love," and its derivatives, studying them in connection with verbs of anger and fear. These emotion verbs are *adāru* ("to be afraid; fear"), *agāgu* ("to become furious"), *ezēzu* ("to become angry; rage"), *galātu* ("to tremble; be afraid"), *kamālu* ("to become angry"), *labābu* ("to rage"), *palāḫu* ("to fear; revere"), *parādu* ("to be scared; be terrified"), *raʾābu* ("to shake; tremble"), *šabāsu* ("to be angry"), *šaḫātu* ("to be afraid; fear; hold in awe"), *šamāru* ("to rage; be furious"), and *zenû* ("to be angry"). The most common derivatives of these verbs are also included in the analysis.

We separated the emotion verbs and their derivatives according to the genre of the document in which they appear. Each emotion word was given a number, according to the genre,

and our methods treat identical lemmas with different numbers as separate words. For example, *râmu*_5 marks the word *râmu* attested in letters, whereas *râmu*_11 marks *râmu* appearing in royal inscriptions. The dataset itself was not divided into genres, and *râmu*_5 and *râmu*_11 are thus analyzed as part of the same dataset of 7,346 texts.

In the full dataset, before the genre division, the word *râmu*, "to love; love," appears 174 times. There are also 11 derivatives of the verb (with the number of occurrences in parentheses), namely *murtâmu* (4, "loving each other; lovers; friends"), *narāmtu* (23, "beloved; favorite," fem.), *narāmu* (134, "loved one; love"), *rāʾimānu* (5, "one who loves"), *raʾīmu* (11, "loved; beloved"), *rāʾimu* (113, "one who loves"), *rāʾimūtu* (1, "friendship"), *rāmu* (6, "loved; beloved"), *rîmu* (1, "beloved"), *ruʾāmu* (13, "love; allure; lovemaking"), and *rûmtu* (3, "beloved," fem.). We only included the derivatives that appeared in the dataset at least nine times before assigning them different genre numbers. After the genre division, we excluded the emotion words which were attested less than five times in a given genre. The same procedure was applied to *râmu* and its derivatives and to the emotion words related to anger and fear. To arrive at our results with minimal preprocessing of the data, we only distinguished between homonymous lemmas that were identical to the aforementioned emotion verbs or their derivatives. We left other homonyms in the dataset unaltered. Homonyms that derived from the same verb and were semantically related but belonged to different parts of speech were not distinguished from each other. As a result, the lemma *râmu*_(number) includes the verb "to love" and the noun "love" but not the homonymous verb "to present to; endow."

All our research data, including texts, statistics, PMI and fastText results, and linguistic networks, is openly available online at https://doi.org/10.5281/zenodo.5861579. Only some aspects of the rich dataset are analyzed in this article, and we hope that other scholars will utilize it in their own research. Since the text material in Oracc is constantly being updated and expanded, its current contents and the numbering of texts may differ from the data available to us in February 2019. The Oracc texts and their metadata used in our research can be accessed via our online repository, but the links in this article point to the current versions of the texts in Oracc.

Methods

Word Similarity

Word similarity is measured by looking at the tendency of words to appear in the same or similar contexts or both (Chandler 2007, 83–88; Levy et al. 2015, 216). The methods used in this chapter, pointwise mutual information and the fastText implementation of the continuous skip-gram model, measure similarity on two levels. PMI can measure the first-order similarity, that is, words appearing in the context of each other and thus belonging to the same *contextual semantic domain*. The nature of the associations depends largely on the used window size, which defines the maximum allowed distance between the observed words. Small windows tend to find compound words and typical attributes associated with other words ("bank transfer;" "ocean floor"), whereas large windows capture more abstract semantic connections ("bank" ~ "mortgage," "loan," "money;" "ocean" ~ "fishing," "algae," "cruise"). These connections are called *syntagmatic* relationships.

In addition to the first-order similarity, fastText also measures the second-order similarity of words that can be found in similar contexts, although not always appearing together. Such words are interchangeable and have a similar semantic function, belonging to the same *lexical semantic domain* and sharing *paradigmatic* relationships. From this perspective, the word "to

speak" is similar to "to talk," "to mumble," and "to whisper." In our previous studies, we have noticed that PMI and fastText do give similar results in our data (Svärd et al. 2018; Svärd et al. 2021a). However, with fastText, we can add to our analysis some words with second-order similarity, that is, words that do not co-occur together but can be used in similar contexts. For example, fastText results indicate that the words *rāʾimu*, "one who loves," and *raʾīmu*, "loved one," appear in similar contexts in literary text genres although they do not occur together in these texts (see "Literary Text Genres").

Pointwise Mutual Information With Context Similarity Weighting

Pointwise mutual information is a word association measure used in automatic collocation extraction (Church and Hanks 1990). The underlying idea of PMI is to represent the statistical association of two words as a ratio of their observed co-occurrence probability to the expected chance of their independent co-occurrence. In mathematical terms, this can be described as follows:

$$PMI(a;b) = \log_2 \frac{\rho(a,b)}{\rho(a)\rho(b)} \tag{1}$$

Theoretically, this formula equals comparing a real-world corpus with its copy – one in which the word order has been infinitely randomized and all the syntactic and semantic constraints of the language are lost. The hypothesis is that if two words show similar co-occurrence patterns in both of these environments, it is unlikely that they bear any meaningful association with each other (Church and Hanks 1990). PMI indicates words that co-occur independently (or more rarely) by giving them a score of 0 or below. If the words are given a positive score, they may be considered collocates.

As PMI has a well-known tendency to give high association scores for low-frequency words, we chose to use a measure called PPMI^2 (Role and Nadif 2011), based on earlier work by Daille (1994):[5]

$$PPMI^2(a;b) = \frac{\rho(a,b)^2}{\rho(a)\rho(b)} \tag{2}$$

PPMI^2 scores are distributed between 0 and 1. A score of 0 indicates that the words are never found within the same window, whereas 1 indicates that the words are perfectly associated and that they co-occur only within the given window. As perfect dependence between words is very rare, especially when larger window sizes are used, the scores tend to be rather small. An example of perfect dependency in an English corpus would be a city name such as *Kuala Lumpur*, as it is very unlikely that either of the words occur elsewhere than adjacent to each other. In our dataset, it is typical that the best co-occurrences receive a score of 0.1 or less.[6]

In order to reduce the impact of repetition in Akkadian texts, we penalize the PPMI^2 scores with a context similarity weight φ, which measures the degree of contextual similarity of the words in question (Sahala and Lindén 2020). This weighting mechanism reduces the significance of word co-occurrences that do not convey previously unseen information due to full or partial repetition or duplication in texts. The weight φ is defined as the average of relative frequencies of unique context words in each position of the collocational window. For example, if all the contexts for PPMI^2 $(a;b)$ are exactly the same, we multiply the score by $\frac{1}{N}$, where

Table 3.1 An example of two word co-occurrences with the context similarity weight $\varphi \approx 0.67$, reducing the final PMI score by ⅓

Word position within window	*1*	*2*	*3*	*4*	*5*
Co-occurrence 1	*a*	*x*	*y*	*w*	*b*
Co-occurrence 2	*a*	*z*	*y*	*w*	*b*
Proportion of unique words	–	1	½	½	–

N equals the co-occurrence frequency $f(a;b)$.[7] Similarly, if two of the contexts are unique, the weight is $\frac{2}{N}$, taking only two co-occurrences into account. This penalty is also applied to partially similar contexts, following the same principle: in the case of only one unique context and another context with one third of the context words being unique, we would multiply the final score by a context similarity weight of $\frac{1+\frac{1}{3}}{N}$. An example is presented in Table 3.1.

Here words *a* and *b* co-occur twice, and their context is defined by window positions 2–4 consisting of words *w*, *x*, *y*, *z*. The proportion of unique context words would then be $\frac{2}{2}$ for position 2 and $\frac{1}{2}$ for positions 3 and 4. We may now calculate the context similarity weight as an average of the proportions of unique words. Note that words *a* and *b* are always ignored from the counts to avoid all scores being penalized. For example, a position containing words [*z*, *z*, *b*] would have a proportion of unique words of ½, as if it only consisted of occurrences of *z*.[8]

Continuous Skip-Gram Model Implemented Using FastText

The so-called continuous skip-gram model (Mikolov et al. 2013b) with negative sampling (Mikolov et al. 2013a) is one of the neural-network-based methods for understanding word similarity (Levy et al. 2015). The continuous skip-gram is a predictive method where the weights of the word vectors predict the contexts in which the words appear (Baroni et al. 2014; Levy et al. 2015). Such predictive methods create word vectors where each word is represented by numbers that indicate its place, and similar words tend to group near each other in the vector space. Two word vectors can be compared with each other by calculating the cosine of the angle between the vectors, ranging between −1 and 1. If the cosine – usually called "cosine similarity" when used to compare word vectors – is close to 1, the angle between the vectors is very small, and they point almost in the same direction. This indicates that the words represented by the vectors appear in very similar contexts (Jurafsky and Martin 2019). The negative sampling speeds up the creation of word vectors, as the weights of all the other words are not adjusted for each word handled but instead a sample of words is taken and their weights updated.

We use a tool called fastText (Bojanowski et al. 2017) to implement the continuous skip-gram model with negative sampling (see Jauhiainen and Alstola Forthcoming). The developers of fastText refined the model and added the capacity to take the subword information into account by representing a word as sequences of characters derived from that word (Bojanowski

et al. 2017). The continuous skip-gram model was developed to handle large corpora of billions of words faster than methods like PMI, which count the co-occurrences of words instead of predicting the contexts (Levy et al. 2015). It has been suggested that the predictive methods are better at capturing word similarity (Baroni et al. 2014), but Levy et al. (2015) have shown that the count-based methods can be tweaked to achieve the same performance by optimizing the hyperparameters in the same way they are used in the predictive methods. In our future research, we are planning to experiment with PMI or some other count-based method to create word vectors in our relatively small Akkadian dataset. Moreover, as fastText analyzes words as sequences of characters, future investigation should focus on how our choice of suffixing emotion words with genre numbers (e.g., *râmu*_5 and *râmu*_7) affects the results.

Network Analysis and Visualization

PMI and fastText provide their results as lists of the best collocates for and the most similar words to the target word, including the respective PMI scores and cosine similarities. Such lists can be laborious to analyze, especially when several target words are compared with one another. An efficient way to study relations between multiple words is to conceptualize and analyze them as networks (Cong and Liu 2014; Quispe et al. 2021). Networks or graphs consist of *nodes* which are connected to each other by *edges* (Newman 2018). Both nodes and edges can be given attributes; these can be used to indicate, for example, edge weight, that is, the strength of the connection between two nodes. Networks can be analyzed without taking edge weights into account, but in most cases, it makes a significant difference whether two people share a strong tie of family relationship or a weak tie of acquaintanceship (in a social network) or whether two words co-occur rarely or frequently (in a linguistic network).

In networks of words, nodes represent words and edges represent relationships between the words. Words that appear together within a given window are connected to each other in a co-occurrence network (Cong and Liu 2014). The edge weight in co-occurrence networks can indicate the simple number of co-occurrences between two words or it can be calculated using a more nuanced formula. We create a co-occurrence network using PMI scores as edge weights. Co-occurrence networks can capture syntagmatic relationships between words but they do not contain information on paradigmatic relationships (see "Word Similarity"; Quispe et al. 2021). Therefore, we also create a network of paradigmatic relationships, using cosine similarities from fastText as edge weights. Figure 3.1 provides examples of the networks

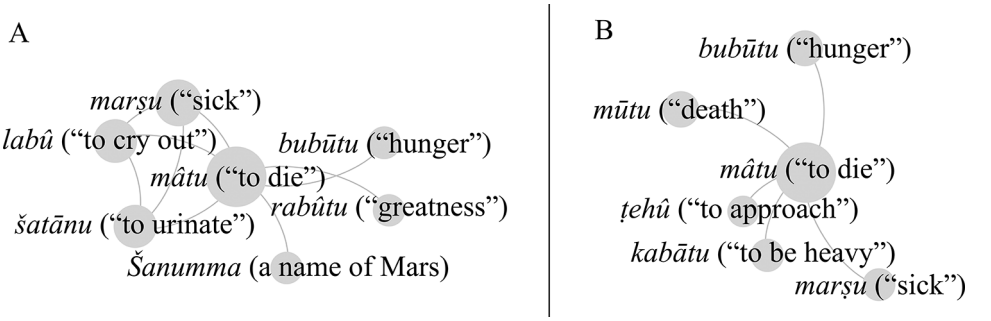

Figure 3.1 Examples of linguistic networks created using PMI scores (A) and cosine similarities from fastText (B) as edge weights.

Source: The data derives from Jauhiainen et al. 2021.

created by using PMI scores and cosine similarities from fastText. The networks display Akkadian words related to the verb *mâtu*, "to die." The network created with PMI (A) displays syntagmatic relationships between *mâtu* and words related to physiological symptoms and astrological observations related to impending death. The fastText network (B) highlights paradigmatic relationships between the words "to die" and "death" and two verbs (*ṭeḫû* and *kabātu*) that refer to approaching or serious illness. Note the overlap between the PMI and fastText results, which is a frequent phenomenon in our relatively small Akkadian dataset.

We use Gephi software (Bastian et al. 2009) to build and visualize our networks and analyze their structures. In addition to the graphs presented in this chapter, full networks and network data are available in our online repository at https://doi.org/10.5281/zenodo.5861579.

Corpus Search Tool Korp

Korp is a web-based concordance tool that allows users to make queries about words in text corpora, one or several corpora at a time. Since the instances of the words searched for are listed with the surrounding words, Korp is a useful tool for studying the contexts in which words appear. Korp software was originally developed by the Language Bank of Sweden (Borin et al. 2012), but there are several online services in different countries that offer their own corpora. The Korp service provided by the Language Bank of Finland[9] contains the Oracc in Korp corpus (Jauhiainen et al. 2019). Since the data in Oracc is constantly changing, the Oracc in Korp corpus is also updated from time to time. At the time of writing, the data in Oracc in Korp was last downloaded from Oracc in May 2019. Most of the texts used for our current dataset are present in that version.

Each cuneiform document present in Oracc in Korp contains metadata such as genre, period, language, and so on. The entry for each word contains metadata about transcription, lemma, translation, part-of-speech tag, and so on, provided this information is available in Oracc. There are three search options in Korp. With the simple search option, it is possible to search with the surface form of the word. In the case of the cuneiform texts in Oracc in Korp, this is the transliteration of the word – that is, the representation of the signs in Latin script. With the extended search option in Korp, one can search by metadata information, such as the lemma or genre, as well as by several consecutive words. Different search criteria can be combined to narrow or enlarge the query. There is also an advanced search option which allows the use of *corpus query processor* (CQP) query language to make even more complicated searches. Korp also supports collection of statistics regarding the metadata categories, such as the distribution of the text genres in which a certain word is attested. Each word in the search results also contains a link to the original text in Oracc.

With Korp, we study individual words or co-occurrences of two words within a certain distance from each other, and we can do this even when there are hundreds or thousands of instances. By studying the contexts in which words are used together, we can determine why our digital methods indicate a close connection between those words. With the help of Korp, we can conclude whether the connection we see relates to the semantic similarity of the words or some specific peculiarity of the text, such as a recurrent list of specific words or phrases.

Workflow

We have created a workflow for using digital tools to study the semantic contexts in which the words of our dataset occur and how these words relate to one another. To summarize our methodology, we study the semantic contexts of a word by using lists of PMI and fastText

results, visualizations in Gephi, and contextual information in Korp. By using the graphs, raw lists, and Korp – and going back and forth between these three tools – we can study words and their semantic contexts very effectively.

Implementing PMI with the Pmizer tool (Sahala 2019), we find the collocates of our words of interest and create lists of the ten best collocates and their PMI scores for each word of interest. Similarly, we use fastText to create word vectors for all the words in our dataset, and using cosine similarity, we calculate which vectors are most similar to the vectors of our words of interest. We then create lists of the ten most similar words and their cosine similarities. The word of interest itself is omitted from these lists. We use a symmetric window of ten words, meaning that PMI and fastText analyze the word of interest in relation to the nine words that immediately precede and the nine words that immediately follow the word of interest in the text. PMI always uses the specified window size of ten words, but fastText randomly selects a window size of 1–10 words on each occasion.

To visualize our results, we import the lists generated with PMI and fastText into Gephi (for a similar approach, see Elwert and Gerhards 2017). The PMI results are presented in one graph and fastText results in another graph. Each word is presented by a node, and edges are created between a word of interest and the ten words appearing in its list. The weight of an edge between two nodes is either their PMI score or cosine similarity. In case two words of interest appear on each other's lists, we take the average of their PMI scores or cosine similarities when assigning the edge weight. For visualization, we run the ForceAtlas2 layout algorithm to position the nodes in relation to each other (Jacomy et al. 2014). These choices are based on our previous research and experimentation with the data and various visualization algorithms in Gephi (Alstola et al. 2019; Svärd et al. 2021a). As our data and the full networks are freely available online (https://doi.org/10.5281/zenodo.5861579), they can be analyzed by other research teams using different layout algorithms and visualization software.

Gephi provides the most efficient and convenient point of departure for analysis, as one can see at a glance not only the immediate neighbors of a word but also its wider neighborhood. This allows us to perceive patterns which would be difficult to recognize by looking at the lists alone (Svärd et al. 2018). In the case of emotion words, for example, it is interesting to see how certain emotion words are clustered together while other emotion words do not have even an indirect connection to that group. At the same time, we use the lists of PMI scores and cosine similarities to evaluate the picture emerging from the graphs, as the raw numbers provide the most accurate view of the best collocates and most similar words.

Finally, we use Korp to study the words of interest in their larger context and access full texts in Oracc, although the data in Korp is not exactly the same as our data (see "Dataset for This Article" and "Corpus Search Tool Korp"). Korp can be used to find a word or several co-occurring words in a certain text corpus or genre, and thus the search can be adjusted to the needs of the research question at hand. The philological analysis of a word in its full textual context adds to the information obtained with other tools and also raises new questions, which can be studied in light of the PMI and fastText results.

Case Study of *Râmu*

Introduction

To showcase the use of language-technological methods in the study of emotion words, we chose to analyze the verb *râmu*, "to love," and its derivatives in our dataset of 7,346 Akkadian texts. According to the main source for lexical semantics in Akkadian, the *Chicago Assyrian Dictionary*

(CAD R, 137–45), *râmu* has the following main meanings: to love one another (as an emotional relationship), sexual attraction, to caress each other, to be loyal to one's earthly or divine overlord, to cherish a dependent or a favorite place, to love a prayer or a virtue, and to have a preference for something. Based on the CAD examples, *râmu* also seems to form an antithetical pair with *zêru*, "to dislike; hate." There are other Akkadian words connected with love, affection, and lovemaking (e.g., *ḫâbu*, *menû*, and *šūdadu*; see Jaques 2006, 129–31), but because of its density and frequency, *râmu* is especially well suited for a study with our methods. As the verb is widely attested and has several common derivatives, we wanted to explore how its usage and semantics vary between different textual genres and in which contexts its derivatives are used.

The ancient Mesopotamian texts themselves are rarely explicit about the genre to which they belong. Applying genre labels based on a modern understanding of literature is not ideal, but following the lead of Benjamin R. Foster, here we consider these genre labels helpful tools rather than comprehensive categories (Foster 2007, 3; Halton and Svärd 2018, 29–30). Moreover, the genre labels assigned to the texts used in this article wholly originate from the work done in Oracc projects. We have merely grouped them according to content and form. This resulted in 14 different genres (or groups of genres) where *râmu* could be examined.[10] Some of these genres can be easily analyzed by a traditional close reading of texts, but some are better approached with digital methods.

In our dataset, *râmu* and its derivatives appear 441 times: 260 times in royal inscriptions, 95 times in literary text genres, 31 times in letters, and fewer than 22 times in each of the other genres. These numbers only include the words to which we assigned a genre number and which are attested at least five times in a given genre. While this distribution is partially due to the number of available texts in each genre in Oracc, it also seems to reflect the actual usage of *râmu*. Although Neo-Assyrian letters, legal compositions, and other text types edited in the State Archives of Assyria series are well represented in our dataset, *râmu* and its derivatives are rarely attested in these texts. Here we have chosen the three genres with the most attestations of *râmu* and its derivatives – royal inscriptions, literary texts, and letters – for a closer look. The letters were chosen as a topic even though *râmu* is not well represented in them, as letters as a genre provide an everyday counterpoint to the more official genres of royal inscriptions and literary texts. The choice of these three genres is also based on the availability of recent philological studies against which we can compare our results. There is extensive scholarship on love literature (see Wasserman 2016 and the works cited there), and the usage of *râmu* has been recently studied in royal inscriptions (Bach, this volume) and letters (Podany, this volume) as well. Although in the following we can only analyze certain aspects of our rich dataset, our entire research data, full networks, and the PMI and fastText results from other genre groups are available in our online repository (Alstola et al. 2022) at https://doi.org/10.5281/zenodo.5861579.

Royal Inscriptions

The word *râmu* and its four derivatives appear 260 times in royal inscriptions in our dataset (the words are marked with the suffix "11" in the graphs and dataset; e.g., *râmu*_11). A look at the graph created with fastText shows that the words *râmu*, "to love; love"; *rāʾimu*, "one who loves"; *narāmu*, "loved one; love"; and *narāmtu*, "beloved; favorite (fem.)" are clustered closely together. The word *ruʾāmu*, "love; allure; lovemaking," is further away and unconnected to the cluster of the four other words (Figure 3.2). These five words are much more scattered in the PMI network, with only *râmu* and *rāʾimu* sharing a common collocate, *šangûtu*, "priesthood" (Figure 3.3). The networks indicate that *râmu*, *rāʾimu*, *narāmu*, and *narāmtu* are used in similar

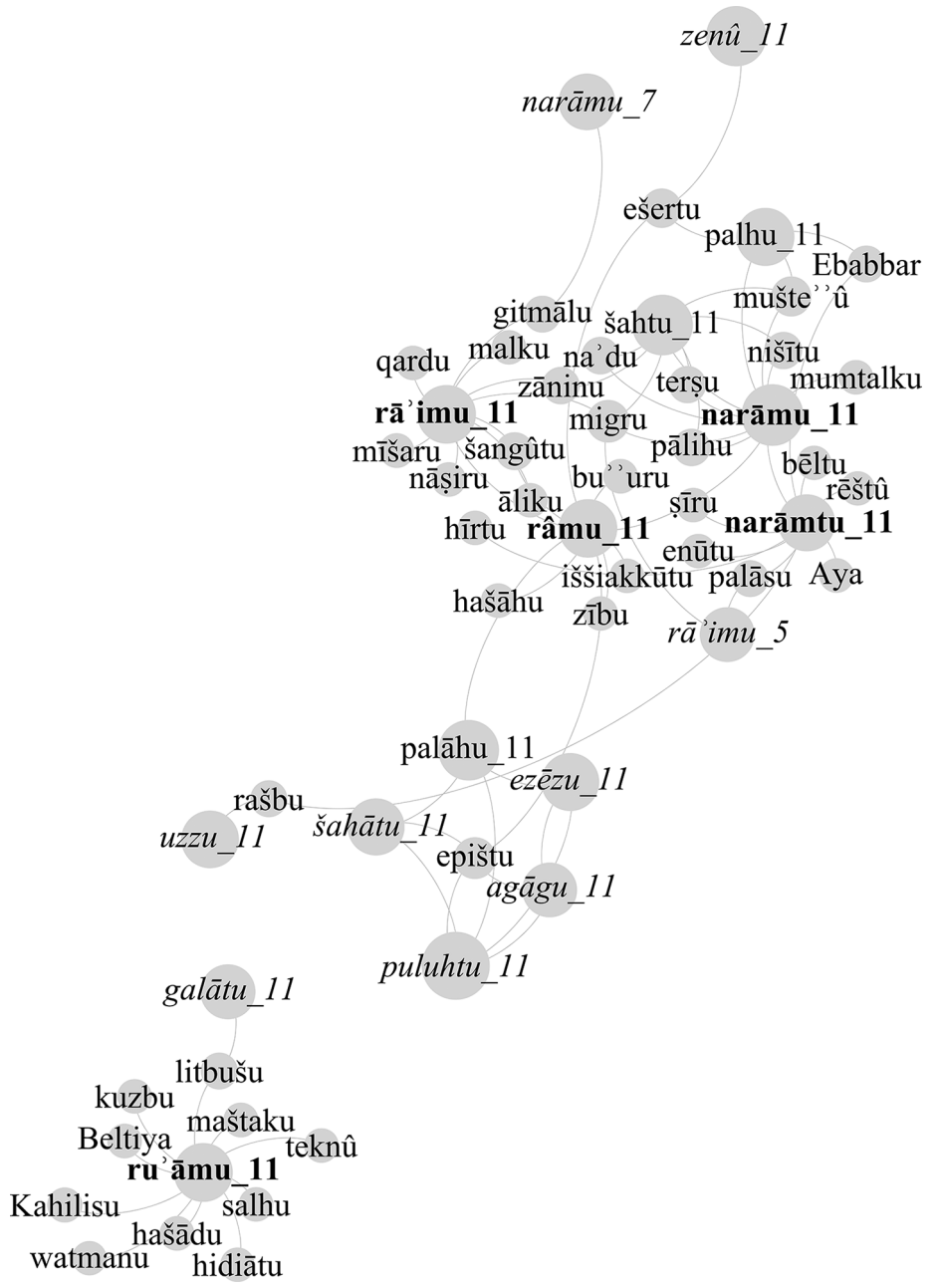

Figure 3.2 Ego networks of *râmu* and its four derivatives in the genre of royal inscriptions, created using cosine similarities from fastText as edge weights.

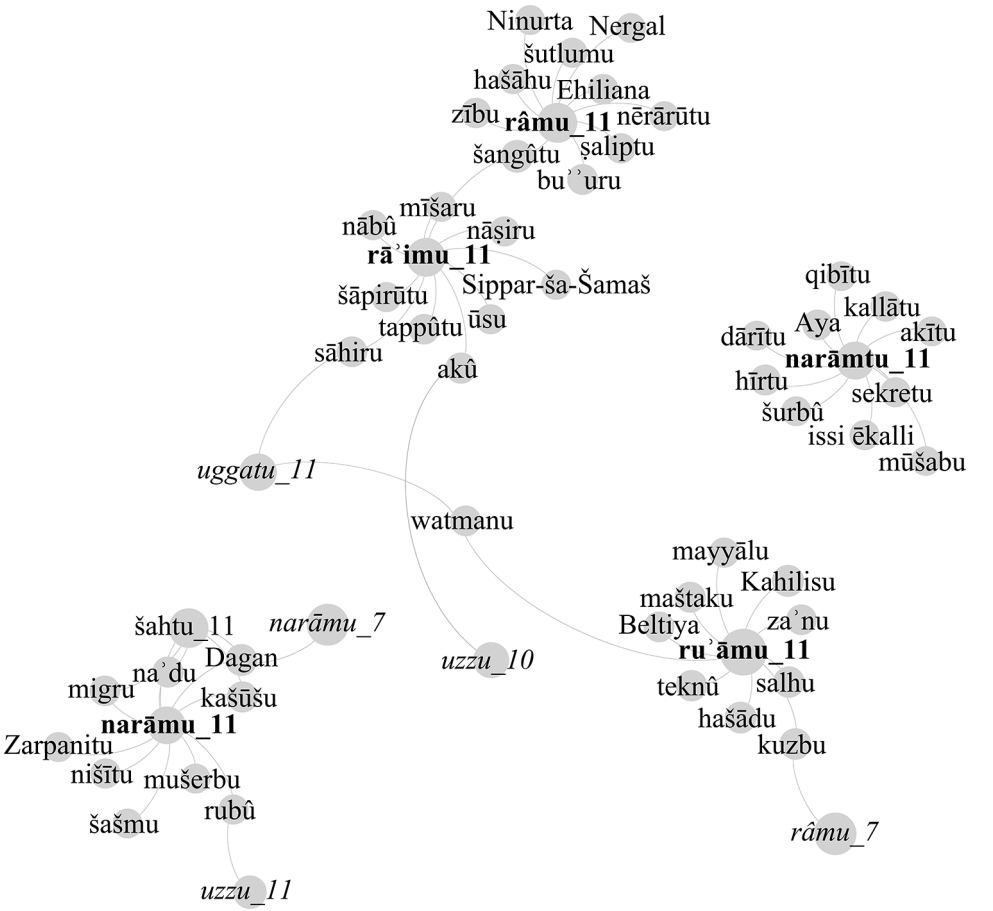

Figure 3.3 Ego networks of *râmu* and its four derivatives in the genre of royal inscriptions, created using PMI scores as edge weights.

contexts and share paradigmatic relationships, but these words do not occur with one another or with the same collocates. At the same time, *ruʾāmu* seems to have its own distinct context of use. Figures 3.2–3.7 depict the PMI and fastText networks of *râmu* and its derivatives in different genres. In addition to the words directly connected to the target words (an ego network at a depth of 1), indirectly connected emotion words are also included in the networks (words from an ego network at a depth of 2). The target words – *râmu* and its derivatives – are in bold, and the indirectly connected emotion words are in italics. All edges have a thickness of 1.

According to the PMI results, the contexts of the verb/noun *râmu* (62 attestations) primarily relate to the relationship between the king and the gods. The subject of the verb is often a divine being who is content with the correct behavior of the king and endorses the king's priestly service and earthly rule. This also explains the paradigmatic relationship between *râmu* and the words *narāmu*, "loved one; love," *rāʾimu*, "one who loves," and *palāḫu*, "to fear; revere," indicated by the fastText results. Both *narāmu* and *rāʾimu* are used to express the gods' love of the king and his actions (see below), and *palāḫu* denotes a reciprocal action, the king's reverence for the gods (Svärd et al. 2021a, 487–89).

In Neo-Assyrian royal inscriptions, the gods are often said to love the priesthood (*šangûtu*) of the king, and the word *šangûtu* becomes the best PMI collocate of and most similar fastText word vector to *râmu* (e.g., RINAP 1 Tiglath-pileser III 47: 12, http://oracc.org/rinap/Q003460/). In another variant of this expression, the gods desire (*ḫašāḫu*) the king's priesthood or priestly services and love his offerings (*zību*) to the gods (e.g., RINAP 5 Ashurbanipal 7 i 80′, http://oracc.org/rinap/Q003706/). The words *šangûtu*, *ḫašāḫu*, and *zību* rank within the ten best PMI collocates of *râmu*, which is not surprising given the clear syntagmatic relationships between these words. However, the same words can be found within the top ten results of fastText, although there is an evident paradigmatic relationship only between *ḫašāḫu* and *râmu*. As we have shown in our previous research, the results of fastText and similar methods may resemble the PMI results in small, repetitive Akkadian datasets, and they do not always show as clear paradigmatic relationships as expected (Svärd et al. 2018; Svärd et al. 2021a).

In a number of Assyrian royal inscriptions, the divine love of the king's priesthood is connected to the motif of hunting (*buʾʾuru*). The gods Ninurta and Nergal, who love the king's priesthood, grant (*šutlumu*) animals to the Assyrian king and command him to hunt (e.g., RIMA 2 A.0.101.30: 84–86, http://oracc.org/riao/Q004484/). The king performs the task successfully and kills or captures hundreds of lions, bulls, and other wild animals. There is a relationship between the gods' love of the king's priesthood, the king's obedience to divine commands, and his success in royal duties such as hunting (see Watanabe 2002, 69–72). This context explains the connection between *râmu* and its PMI collocates *Ninurta*, *Nergal*, and *šutlumu*. The word *buʾʾuru* ranks high in both the PMI and fastText results.

Other contexts of the word *râmu* are visible in our PMI results as well. In Esarhaddon's (reigned 680–669 BCE) royal inscriptions, the king designates himself as someone "who loves loyalty (*kittu*) and regards treachery (*ṣaliptu*) as taboo (*ikkibu*)" (e.g., RINAP 4 1 iv 25–26, http://oracc.org/rinap/Q003230/). The statement relates to a rebellion of the Arabs against their vassal king, who was installed by the Assyrians, and it emphasizes Esarhaddon's expectation of loyalty and honesty from his subjects. In yet another context, the goddess Nanaya loves her shrine Ehiliana, to which King Ashurbanipal (reigned 668–631(?) BCE) returned her statue from Elam (e.g., RINAP 5 Ashurbanipal 9 v 72–vi 11, http://oracc.org/rinap/Q003708/).

The word *narāmu* (102 attestations), "loved one; love," is frequently used in Assyrian and Babylonian royal inscriptions in which kings introduce themselves and their epithets (e.g., Ashurnasirpal II, RIMA 2 A.0.101.1 i 9–17, http://oracc.org/riao/Q004455/; Esarhaddon, RINAP 4 104: i 1–18, http://oracc.org/rinap/Q003333/; Nabopolassar, Da Riva 2013 C22 i 1–19, http://oracc.org/ribo/Q005374/). The word usually appears in the expression *narām DN*, "loved one of (a deity)," which is used to indicate the close relationship between the king and the gods and emphasize the divine sanction of rulership. The king is the one who is loved by the gods and who takes care of cultic duties and executes the divine will. This usage of *narāmu* dominates our computational results, and the ten best PMI collocates and the ten most similar words according to fastText belong almost exclusively to the context of royal epithet lists. A few royal epithets rank high in both the PMI and fastText results, including *šaḫtu*, "reverent; humble"; *naʾdu*, "attentive; reverent"; and *migru*, "favorite (of a god)." These epithets have clear paradigmatic associations with *narāmu*, as they all describe an intimate relationship between the king and his gods. At the same time, the relations are also of a syntagmatic nature, because the words co-occur in the context of epithet lists.

The case of *narāmu* is an illustrative example of the strengths and caveats of the computational and statistical approach. On the one hand, PMI and fastText successfully identified the most important usage context of *narāmu* in royal epithets. On the other hand, a closer inspection reveals that the word also occurs in other semantic contexts in royal inscriptions.

In several inscriptions of Sennacherib (reigned 704–681), Nineveh is called "the city loved by (the goddess) Ishtar" (e.g., RINAP 3 1: 63, http://oracc.org/rinap/Q003475/), and Esarhaddon repeatedly calls crown prince Ashurbanipal "my beloved son" (e.g., RINAP 4 93: 25, http://oracc.org/rinap/Q003322/). These contexts can be discerned relatively easily by using the corpus search tool Korp (see "Corpus Search Tool Korp"), because they occur in five or more texts. However, they are not visible in our PMI and fastText results for two reasons. First, *narāmu* occurs most frequently in lists of epithets, tying it to other royal titles and designations. These words are semantically very similar to *narāmu*, and they rightly occupy the top positions in the PMI and fastText results, leaving no space for words occurring in other contexts. Second, because the inscriptions repeat certain passages almost verbatim, our PMI script assigns a penalty to them. There is more variation in the epithet lists, but the phrases "the city loved by Ishtar" and "my beloved son" are part of a larger sequence which is repeated almost verbatim several times in Sennacherib's and Esarhaddon's inscriptions. PMI and fastText succeed well in identifying patterns, but the study of details still requires philological work; this can be efficiently facilitated by Korp.

The PMI and fastText results for the word *rāʾimu*, "one who loves," are dictated by the inscriptions of Sennacherib, whose list of epithets designates him as one who loves justice (*rāʾim mīšari*). Although *rāʾimu* appears 74 times in royal inscriptions, around 20 attestations of *rāʾimu* in Sennacherib's list of epithets (e.g., RINAP 3 1: 1–3, http://oracc.org/rinap/Q003475/) affect our results so much that more than half of the ten best PMI collocates and the most similar words according to fastText originate from this context. At the same time, the word *šangûtu*, "priesthood," ranks high in the PMI and fastText results. Certain deities are said to love (*rāʾimu*) the priesthood of the Assyrian king in several royal inscriptions which cover a period of four centuries, from Tiglath-pileser I (reigned 1114–1076) (RIMA 2 A.0.87.1: vii 71–73, http://oracc.org/riao/Q005926/) to Esarhaddon (RINAP 4 1 i 74, http://oracc.org/rinap/Q003230/). Only the gods Ashur and Ishtar are mentioned in these passages in the inscriptions of Sennacherib and Esarhaddon, but Ninurta and Ashur figure most prominently in the earlier inscriptions, which also refer to the gods Adad, Anu, and Nergal.

The PMI and fastText results highlight the most prominent contexts in which *rāʾimu* is used, but a closer inspection in Korp reveals a few contexts that are not clearly indicated by the computational methods. Two Middle Assyrian kings – Tukulti-Ninurta I (reigned ca. 1233–1197 BCE) and Tiglath-pileser I – often boast that the gods are the "ones who love me" (e.g., RIMA 1 A.0.78.23: 92, http://oracc.org/riao/Q005859/). In both Assyrian and Babylonian inscriptions, deities are occasionally said to love *šarrūtu*, "kingship," (e.g., RINBE 2 Nabonidus 24 i 12, http://oracc.org/ribo/Q005421/) or *palû*, "(a king's) reign" (e.g., RINAP 5 Ashurbanipal 12 i 3′, http://oracc.org/rinap/Q003711/). Finally, in the Babylonian inscriptions of King Cyrus of Persia (reigned 559–530 BCE), *rāʾimu* expresses the king's love for the city of Babylon and the Esagil and Ezida temples (Schaudig 2001 K2.1: 23 and K1.1: 1–2, http://oracc.org/ario/Q006653/, http://oracc.org/ario/Q006655/).

Two derivatives of *râmu – narāmtu* (13 attestations), "beloved; favorite (fem.)," and *ruʾāmu* (9 attestations), "love; allure; lovemaking" – are infrequently attested in our dataset on royal inscriptions. Not surprisingly, the PMI and fastText connect *narāmtu* to other female words, such as *ḫīrtu*, "(equal-ranking) wife" (PMI and fastText); *kallātu*, "daughter-in-law; bride" (PMI); *issi ēkalli* (MUNUS-E_2.GAL), "queen" (PMI); *sekretu*, "enclosed (woman)" (PMI); and *bēltu*, "lady" (fastText). These words define well the contexts in which *narāmtu* is used, where it refers to the divine spouses of gods and to the wives and daughters of earthly kings. Among them are the goddess Aya (RINBE 2 Nabonidus 16 i 50–51, http://oracc.org/ribo/Q005413/), Sennacherib's queen Tashmetu-sharrat (turn of the eighth and seventh centuries BCE) (RINAP 3 40:

44″, http://oracc.org/rinap/Q003514/), and Nabonidus's daughter En-nigaldi-Nanna (mid-sixth century) (RINBE 2 Nabonidus 34: ii 40, http://oracc.org/ribo/Q005431/). While *narāmtu* is used in both Assyrian and Babylonian royal inscriptions, the derivative *ruʾāmu* is exclusively used by the Assyrians. The PMI and fastText results are very similar, and they both connect *ruʾāmu* to sexual love and the divine world. First, *ruʾāmu* refers to the lovely or alluring shrine (*atmanu ša ruʾāme*) of a deity (RIMA 2 A.0.101.31: 17, http://oracc.org/riao/Q004485/). Second, it denotes a palace of love or lovemaking (*ēkal ruʾāme*), which Sennacherib built for his queen Tashmetu-sharrat (RINAP 3 Sennacherib 40: 44″–46″, http://oracc.org/rinap/Q003514/), and a bed of love-making (*mayyāl taknê . . . epēš ruʾāme*) for the god Marduk and his wife Zarpanitu (e.g., RINAP 5 Ashurbanipal 10: i 46–54, http://oracc.org/rinap/Q003709/).

In general, *râmu*, *narāmu*, and *rāʾimu* do not have an erotic or sexual connotation in royal inscriptions; instead, they primarily describe the relationship between the king and the gods. The gods love the king, who ensures the correct order of things and fulfills the divine will. The similarity of the three words was already apparent in the fastText graph (Figure 3.2), in which they and the word *narāmtu* are clustered together. However, *narāmtu*, "loved one (fem.)," belongs to this cluster only because it resembles *narāmu*, "loved one (masc.)," both words being used to express affection between family members in some texts of the dataset. A closer analysis reveals that the usage of *narāmtu* and *ruʾāmu* is generally quite different from the three other words. They mainly express the relationship between a divine or earthly husband and wife and the sexual love between the spouses. PMI and fastText succeeded in identifying the most typical contexts in which each of the words is used, but Korp was needed to detect some of their rarer usages. The combined digital workflow with PMI, fastText, and Korp yielded very similar results to Bach's (this volume) philological analysis of love words in royal inscriptions.

Literary Text Genres

In the Oracc metadata, a variety of subgenres have been grouped under the genre designations "literary" and "literary work." As the difference between these two designations is slight, we have combined them into a single group. These texts do not constitute a single, well-defined genre such as royal inscriptions or letters, but we rely on the Oracc metadata and refrain from introducing subtler genre divisions into the dataset.

The word *râmu* and its three derivatives are attested 95 times in literary text genres (the words are marked with the suffix "7" in the graphs and dataset; e.g., *râmu_7*). Although several subgenres of literature are present in our dataset, the words *râmu*, "to love; love"; *rāʾimu*, "one who loves"; and *raʾīmu*, "loved; beloved," are almost exclusively attested in love literature. The usage of the derivative *narāmu*, "loved one; love," is more varied, and it also appears in mythological texts. The concentration of *râmu* and its derivatives in love literature is visible in the fastText results, showing that *râmu*, *rāʾimu*, and *raʾīmu* have similar word vectors, which means that they are attested in semantically similar contexts and share paradigmatic relationships. At the same time, these word vectors do not resemble the word vector of *narāmu*. This pattern is quite visible in the graph created using the fastText results, in which *râmu*, *rāʾimu*, and *raʾīmu* are clustered together, while *narāmu* is not directly connected to them (Figure 3.4). This suggests that the texts belonging to the corpus of love literature in Oracc form a semantically coherent group, in which *râmu*, *rāʾimu*, and *raʾīmu* are used quite similarly. Attested in both love literature and mythological texts, the word *narāmu* is used differently.

However, the PMI results and the corresponding graph (Figure 3.5) show that the actual patterns of co-occurrence are more complicated. The words *râmu* and *narāmu* typically co-occur with *rāʾimu* and share common collocates with it. This results in the clustering of *râmu*, *rāʾimu*,

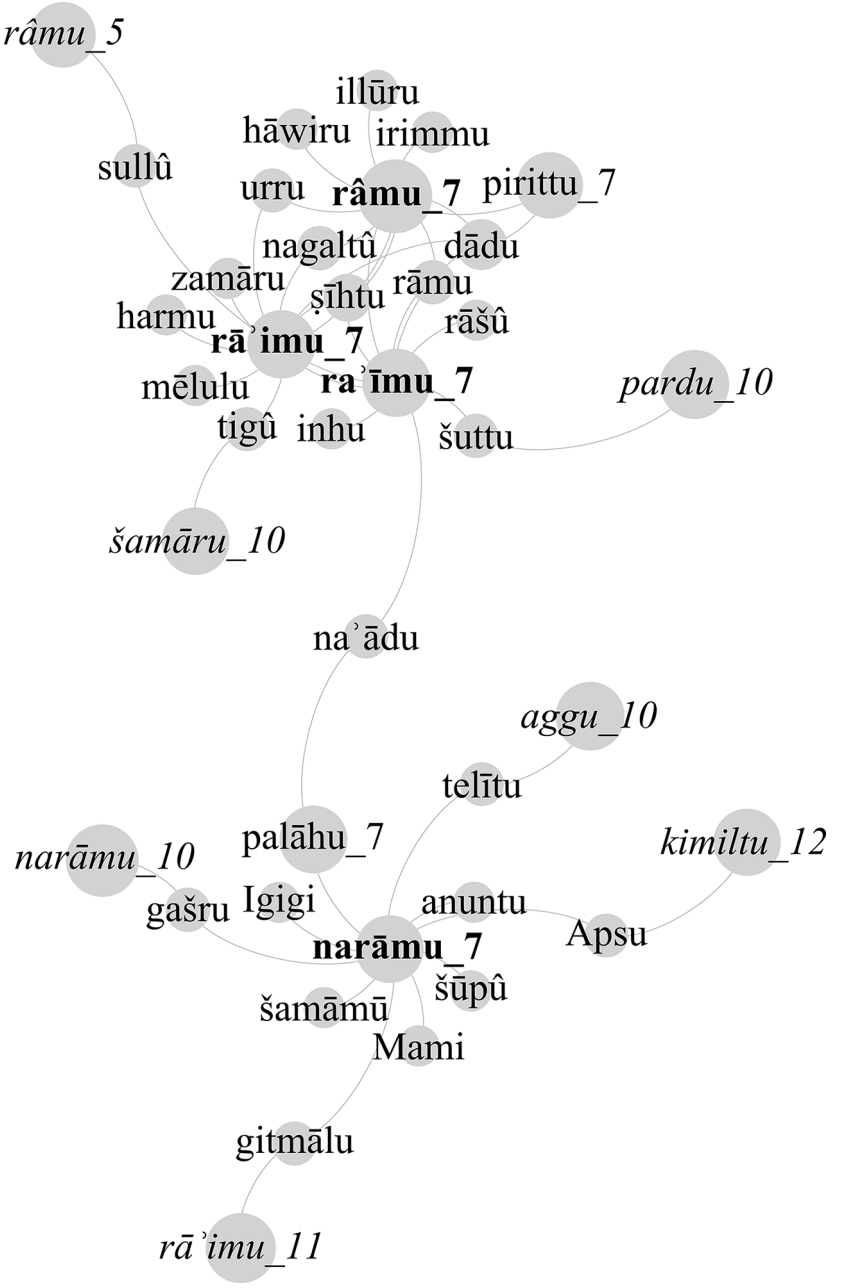

Figure 3.4 Ego networks of *râmu* and its three derivatives in literary text genres, created using cosine similarities from fastText as edge weights.

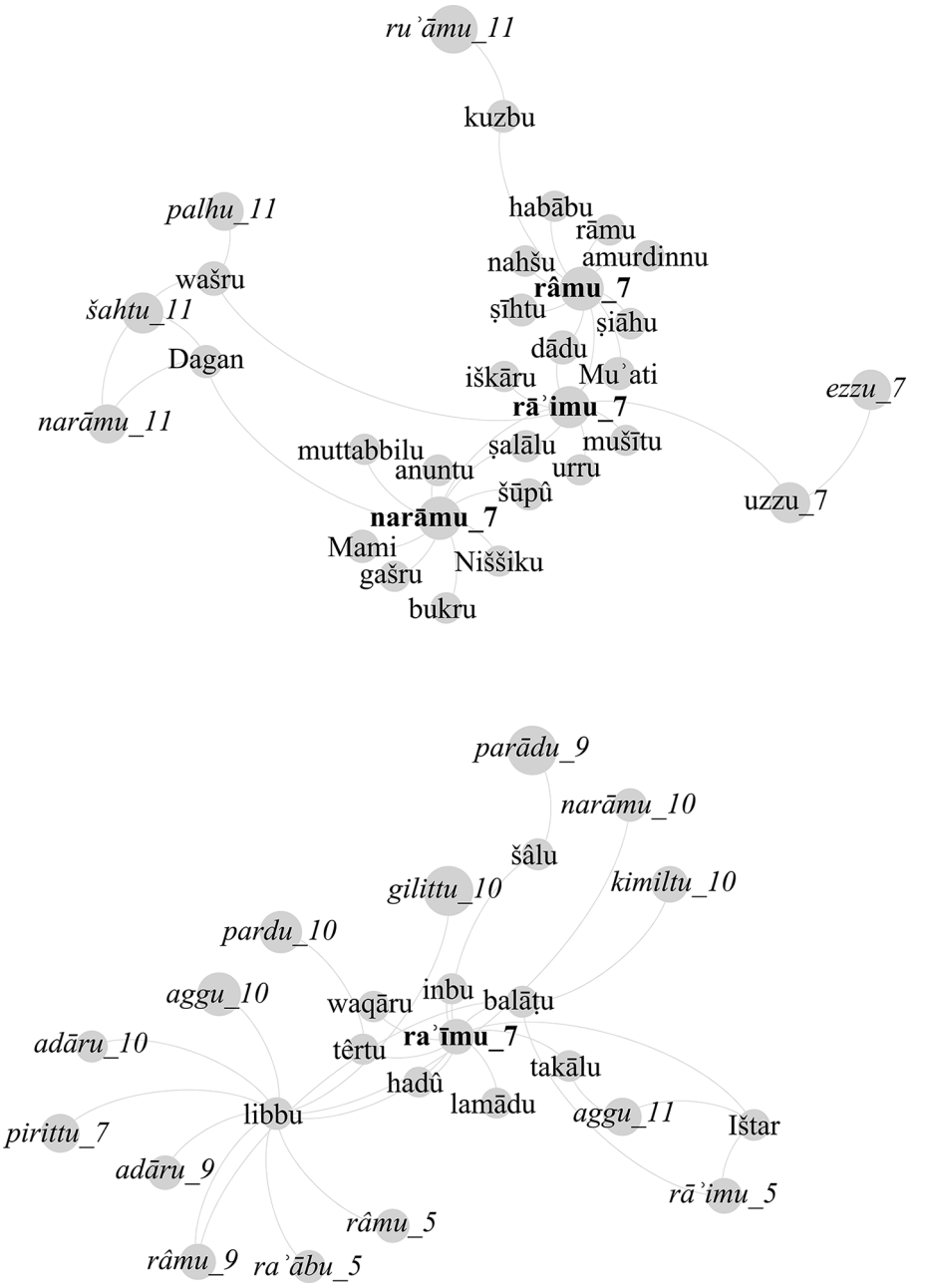

Figure 3.5 Ego networks of *râmu* and its three derivatives in literary text genres, created using PMI scores as edge weights.

and *narāmu* in the graph, although there are no links between *râmu* and *narāmu* or their collocates, showing that they are not typically attested in the same contexts. At the same time, the word *raʾīmu* and its collocates are not connected to this cluster of three words and their collocates. This shows that the similarity between *râmu*, *rāʾimu*, and *raʾīmu* – as detected by fastText – does not result from the fact that all the three words are typically attested together. It is rather the language of love literature in general that makes the contexts of the three words similar.

The Oracc corpus of Akkadian love literature from the third and second millennia BCE originates from Nathan Wasserman's work (2016), which not only contains text editions but also a careful study of the corpus, its themes, and its vocabulary. This enables us to compare the results of our digital analysis with the results of Wasserman's philological study. Wasserman studies the vocabulary of the corpus from the perspective of semantic fields and motifs and metaphors. Our PMI and fastText results are dominated by words relating to Wasserman's semantic field of love and lovemaking, while a few words belonging to Wasserman's field of flora and metaphors of sleep, dreams, and awakening occasionally appear in the results. In general, our results for the most common love word in our dataset, *râmu*, are almost exclusively composed of lexemes that Wasserman also discusses, but the results for the rarer derivatives contain more words that do not feature in Wasserman's analysis.

The verb/noun *râmu* (56 attestations) characterizes erotic and sexual love, and eight of its ten best PMI collocates belong to Wasserman's semantic field of love and lovemaking. According to the PMI results, it co-occurs with words such as *ṣīḫtu*, "laughter"; *dādu*, "darling"; and *ṣiāḫu*, "to laugh" (e.g., Wasserman 2016 no. 2 obv. 6, http://oracc.org/akklove/P251898/; no. 19: rev. vii 41′–45′, http://oracc.org/akklove/P282615/). In this context, *ṣiāḫu* and *ṣīḫtu* refer to sexual joy and lovemaking, and *dādu* in the plural expresses lovemaking or sexual attractiveness (CAD D, 20; CAD Ṣ, 65, 186; Wasserman 2016, 32, 52, 54). The fastText results emphasize how deeply the use of *râmu* and its derivatives characterize the corpus of love literature: there are four derivatives – *rāʾimu*, "one who loves"; *irimmu*, "love charm"; *rāmu*, "loved; beloved"; and *raʾīmu*, "loved; beloved" – among the six most similar words to *râmu*.[11] Although words such as *dādu* and *ṣīḫtu* also appear in the fastText results, fastText makes the paradigmatic relationships between *râmu* and its derivatives more apparent than PMI does. Despite the dominance of the semantic field of love and lovemaking, some words relating to Wasserman's semantic field of flora (*amurdinnu*, a thorny bush; see eSAD) and metaphors of sleep, dreams, and awakening (*urru*, "daytime") are also visible in the PMI and fastText results.

The noun *rāʾimu* (15 attestations), "one who loves," designates both male and female lovers, and it occurs in varying contexts in love literature. Wasserman's semantic field of love and lovemaking is central again; *dādu*, "darling," and derivatives of *râmu* appear among the best PMI collocates. Vocabulary relating to metaphors of sleep and awakening is also attested. The word *ṣalālu*, "to lie (down); sleep" (Wasserman 2016, 45–47), is sometimes used in contradictory ways: sleeplessness is equated to the absence of love (no. 16 obv. ii 6–9, http://oracc.org/akklove/X001013/), but sleep is also something that keeps lovers apart (no. 27–34 obv. 38–40, http://oracc.org/akklove/P355910/). The fastText results corroborate but also nuance the picture given by PMI. The majority of the ten most similar words to *rāʾimu* belong to the semantic field of love and lovemaking or to the metaphors of sleep and awakening. In addition to the many words discussed by Wasserman, we can mention *mēlulu*, "to play," and *nagaltû*, "to awake." At the same time, the fastText results feature words (*zamāru*, "to sing; song," and *tigû*, "(a kind of) drum; (a kind of) song"; also note the PMI collocate *iškāru*, here "song series") that are predominantly used to describe different kinds of songs or hymns in a catalogue of love-related and other literary works (Wasserman 2016, 195–205; no. 19, http://oracc.org/akklove/P282615/).

The derivative *raʾīmu* (9 attestations), "loved; beloved," designates the loved person and is only attested in the corpus of love literature. Despite the small number of occurrences, the fastText results for *raʾīmu* are good. The most similar words include the word *râmu* and its derivatives and words related to Wasserman's semantic field of love and lovemaking (*ṣīḫtu*, "laughter," and *dādu*, "darling") and metaphors of sleep and awakening (*šuttu*, "dream"). The PMI collocates are much more varied; they have relatively low scores, and they are of poorer quality than the fastText results. This probably relates to the rareness of the word *raʾīmu* in literary text genres, but it remains unclear why fastText is not similarly affected by this. The PMI collocates cannot be used to characterize *raʾīmu* in a meaningful way, and only a few of them are referred to in Wasserman's study.

Finally, the derivative *narāmu* (15 attestations), "loved one; love," not only occurs in love literature but also in mythological texts such as the *Anzu Epic* and *Enuma elish* and in the *Gilgamesh Letter*. Our results – the PMI collocates in particular – are dominated by Ninurta's epithet list in the *Anzu Epic*. The word *narāmu* designates Ninurta as the beloved of his mother Mami, and it is used together with epithets like *šūpû*, "resplendent one"; *gašru*, "strong one"; and *bukur Enlil*, "son of Enlil" (e.g., SAACT 3 1: 1–7, http://oracc.org/cams/anzu/Q002769/). Because *narāmu* is used several times in a very specific context in Ninurta's epithet lists but its other occurrences are very varied, the PMI and fastText results do not provide useful information about its other usages in literary text genres. However, a few observations can be made using the corpus search tool Korp (see "Corpus Search Tool Korp"). The word *narāmu* refers to a variety of relationships, designating Marduk as the beloved (son?) of the gods in *Enuma elish* (STT 1 12 rev. 3′, http://oracc.org/cams/gkab/P338328/), Gilgamesh as the beloved of Marduk in the *Gilgamesh Letter* (STT 1 40 obv. 2–3, http://oracc.org/cams/gkab/P338357/), and Dumuzi as the beloved of Ishtar (Wasserman 2016 no. 9 obv. 5–6, http://oracc.org/akklove/P413919/). In love literature, the word also refers to beloved mortal persons (e.g., Wasserman 2016 no. 3 obv. 4–8, http://oracc.org/akklove/P254179/).

In literary text genres, *râmu* and its derivatives *rāʾimu*, "one who loves," and *raʾīmu*, "loved; beloved," are predominantly attested in the corpus of love literature. Their occurrences belong most often to Wasserman's (2016) semantic field of love and lovemaking, but some contexts also feature love-related metaphors of sleep, dreams, and awakening. As our fastText results in Figure 3.4 indicated, these three words are used in similar contexts and they share paradigmatic relationships, whereas the derivative *narāmu*, "loved one; love," is different. It appears both in mythological texts and in love literature, characterizing the loved ones of deities and mortals. The quality of the PMI and fastText results for *râmu* and *rāʾimu* was good, but the repetitive occurrences of *narāmu* in Ninurta's epithet list in the *Anzu Epic* caused difficulties for both methods. Somewhat surprisingly, fastText fared better than PMI in the analysis of the rare derivative *raʾīmu*.

Letters

In the genre of letters (the words are marked with the suffix "5"; e.g., *râmu_5*), the word *râmu* and its derivative *rāʾimu* are infrequent (31 attestations), and the numerical values of both the PMI and fastText results indicate that no strong collocates or very similar words were found. The results are not poor, however, and they successfully highlight the contexts in which *râmu* and *rāʾimu* occur.

In the graph created using the fastText results (Figure 3.6), it is remarkable that *râmu* "to love; love," and *rāʾimu*, "one who loves," remain resolutely apart, unlike in royal inscriptions and literary text genres in which *râmu* and many of its derivatives are clustered neatly together

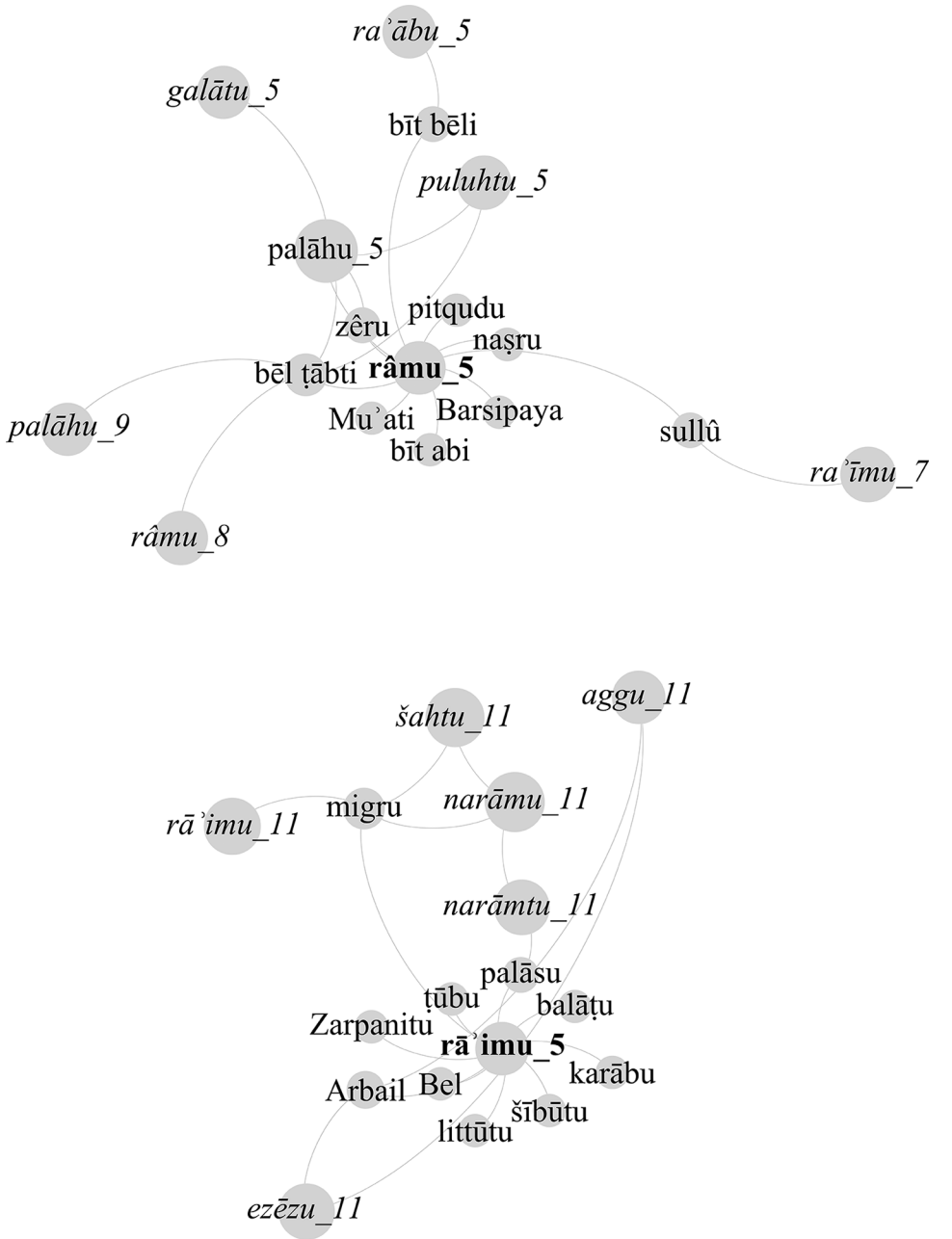

Figure 3.6 Ego networks of *râmu* and its derivative *rāʾimu* in the genre of letters, created using cosine similarities from fastText as edge weights.

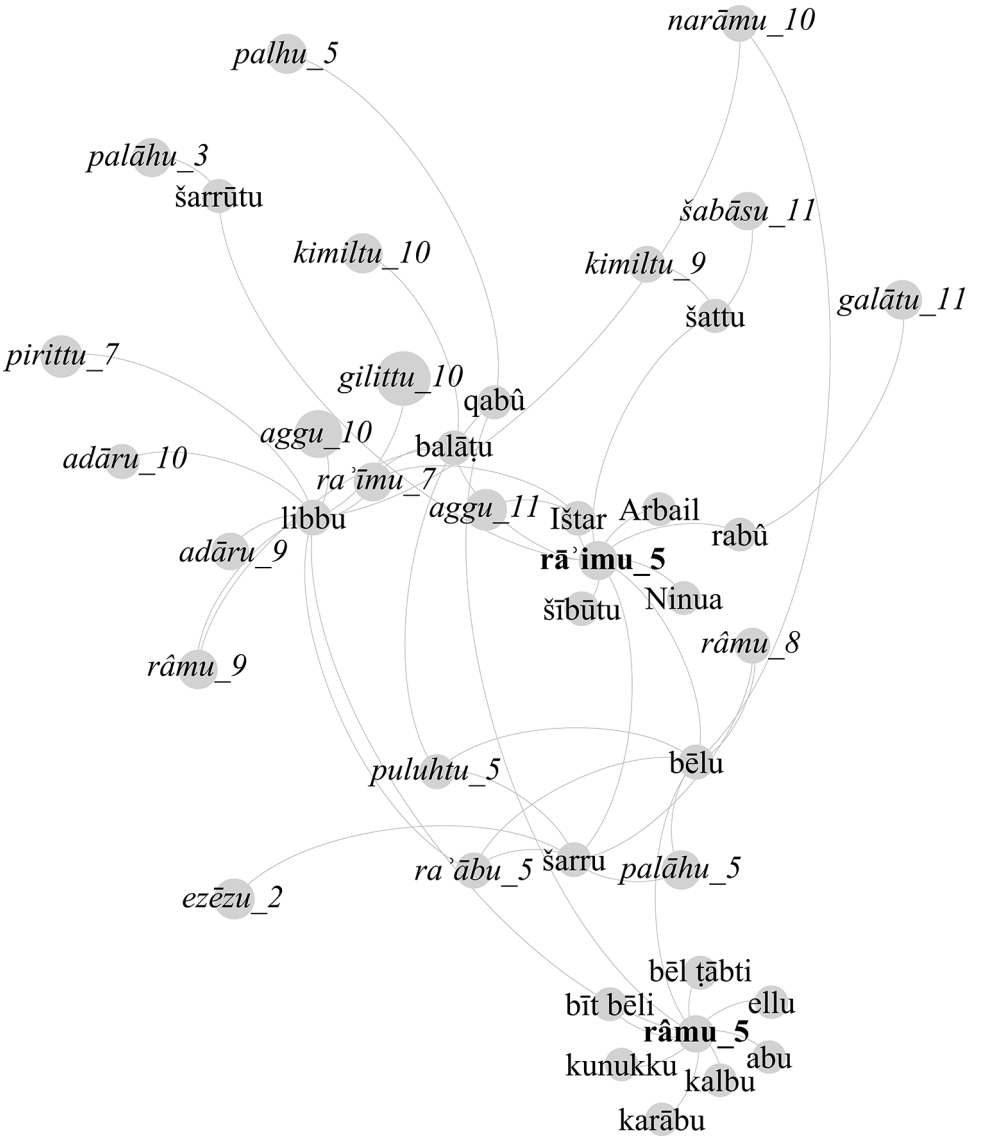

Figure 3.7 Ego networks of *râmu* and its derivative *rāʾimu* in the genre of letters, created using PMI scores as edge weights.

(see Figures 3.2 and 3.4). The word *râmu* is connected to *palāḫu*, "to fear; revere," and it is located in a group of fear and anger words attested in the genre of letters. At the same time, the derivative *rāʾimu* can be found among several emotion words attested in royal inscriptions. The dissimilarity of *râmu* and *rāʾimu* is not surprising, given their very few attestations in the dataset and the repetitive context in which *rāʾimu* primarily occurs (see below). In the graph created using the PMI results (Figure 3.7), however, both *râmu* and *rāʾimu* are located close to other emotion words attested in the genre of letters, including *palāḫu*, *puluḫtu*, "fear; fearsomeness," and *raʾābu*, "to shake; tremble." This clustering is primarily explained by the

words *bēlu*, "lord," and *šarru*, "king," which appear repeatedly in the letters addressed to the king and interlink the previously mentioned emotion words in the graph.

Since the number of occurrences in this genre is small in our dataset, it is worth examining the results in detail in the Korp interface. There, the verb/noun *râmu*, "to love; love," appears in multiple volumes of letters published in the State Archives of Assyria series. The occurrences are fairly evenly spread chronologically and across the letter volumes of the series. The meaning most often expressed relates to appreciation: for example, "(If) you like him for what he is, why am I not intensely loved (*ina libbi tarânšu ina libbi mīni lā urâmuannî*)?" (SAA 1 12: rev. 5–6, http://oracc.org/saao/P334693/; the translations are quoted from the online text editions). But this is appreciation in a very specific sense: it refers to the obligations that a person has toward the object of his love. In other words, it refers to instrumental love, which endeavors to advance the interests of the object of love. Letter SAA 10 198: rev. 9 (http://oracc.org/saao/P334300/) expresses this in an abstract way, asking rhetorically: "Who does not love (*irâm*) his benefactor (*bēl ṭābti*)?" Numerous related occurrences are known: "None of those who serve in the palace like me (*lā iraʾʾumunni*); there is not a single friend of mine (*bēl ṭābtīya*) among them to whom I could give a present, and who would accept it from me and speak for me" (SAA 10 226: rev. 14–19, http://oracc.org/saao/P333954/); "The king, your father, loves (*irâm*) the son of one who worked for him" (SAA 16 34: rev. 14–15, http://oracc.org/saao/P334608/); and "[From] the very beginning I have been a dog who loves (*irâmu*) [the house of] his [lord]" (SAA 18 182: obv. 9–10, http://oracc.org/saao/P237664/). The love of gods toward the king (e.g., SAA 16 105: rev. 12–15, http://oracc.org/saao/P334131/) can be expressed as well.

All these connotations of *râmu* (17 attestations in our dataset), which appear upon an examination of the primary sources, are visible in the PMI and fastText results as well (for example, *bēl ṭābti*, "benefactor; friend"). The PMI results highlight the context and social sphere in which our corpus of letters was written. The collocates refer to hierarchical relationships between the sender and recipient (*kalbu*, "dog," and *bēlu*, "lord"), and they include other vocabulary frequently used in the Neo-Assyrian state correspondence (*qabû*, "to say," and *karābu*, "to bless"). Furthermore, fastText demonstrates its ability to find paradigmatic relationships between semantically similar words, as the antonym *zêru*, "to dislike; hate," and the emotion verb *palāḫu*, "to fear; revere," appear in the fastText results. As observed earlier, *râmu* is part of a group of words related to fear in the fastText network. Its connection is via *palāḫu*, but the clustering algorithm of Gephi clearly places *râmu* in letters as part of a greater "fear-network" (Figure 3.6; compare to Svärd et al. 2021b). The words *naṣru*, "guarded; attentive," and *pitqudu*, "cautious; circumspect," also appear in similar contexts as *râmu*, designating the correct behavior of the king and his subjects. The vocabulary of social relationships is notably present in the fastText results as well (on *bīt bēli*, "domain of the lord"; see Fales 2000; *bīt abi*, "house of the father; paternal estate").

The noun *rāʾimu* (14 attestations), "one who loves," is most prominently attested in letters which the priest Urdu-Nabu sent to King Esarhaddon (SAA 13 56–69; see PNA 3/2, 1408–09). The priest greets the king in most of his letters as follows:

> To the king, my lord: your servant, Urdu-Nabu. Good health to the king, my lord. May Ashur, Sin, Shamash, Marduk, Zarpanitu, Nabu, Tashmetu, Ishtar of Nineveh, and Ishtar of Arbela – these great gods who love your kingship (*rāʾimūte šarrūtīka*) – allow the king, my lord, to live 100 years. May they grant the king, my lord, the satisfaction of old age, extreme old age.
>
> (SAA 13 56: obv. 1–12, http://oracc.org/saao/P334061/)

This expresses the well-known topos of the gods' love toward the king (see "Royal Inscriptions"). All of the ten best PMI collocates and six of the ten most similar words in the fastText results are attested in the greeting formula. These blessings of Urdu-Nabu – and the few additional occurrences – therefore do not shift our interpretation of *râmu* as outlined above but instead strengthen it. The main conclusion regarding *râmu* in letters is that it is used mostly to document obligations between people (and deities). These obligations are often of a hierarchical nature, and *râmu* is something that servants desire, wish for, and want to show to their lord. A similar view emerges from the Amarna letters (fourteenth century BCE); instead of the word *râmu* being generally employed to designate the relationship between two great kings of equal rank, it is used by vassal kings pleading for help and support from the pharaoh, in order to describe the relationship between them and their overlord (Podany, this volume).

Overall, methodologically, the case study of *râmu* in letters strengthens the idea that, in the case of small amounts of material, the analysis of collocation patterns via Korp or a similar interface is faster and more precise than statistical analysis (see also Svärd et al. 2021a). However, the PMI and fastText results were relatively good, and they captured the predominant contexts in which the words of interest occurred: the usage of *râmu* is characterized by vocabulary of obligations and social relations, and the derivative *rāʾimu* is typically used in the context of greetings and blessings. Moreover, statistical analysis was needed to isolate a probable group of texts that could usefully serve as the object of study – something that would have been very cumbersome and slow to do by hand, considering the number of attestations for the verb *râmu* and its derivatives.

Conclusions

Digital methods from the field of language technology allow us to have an aggregate view of large digital text corpora. These methods can be used for a variety of purposes, including the study of individual words and their semantic domains. Emotion words provide a good test case for digital analysis of semantic domains, because a single word can be used in a wide variety of contexts, while several emotion words can have overlapping semantic domains. Although the availability of large and sufficiently digitized text corpora from premodern times is limited, Akkadian sources form a notable exception.

In this chapter, we applied two language-technological methods to study the Akkadian verb *râmu* ("to love"; the lemma also includes the homonymous noun "love") and its derivatives *narāmtu* ("beloved; favorite," fem.), *narāmu* ("loved one; love"), *raʾīmu* ("loved; beloved"), *rāʾimu* ("one who loves"), and *ruʾāmu* ("love; allure; lovemaking"). First, we used pointwise mutual information to detect the collocates that typically appear near our words of interest, highlighting the context in which the target words appear. Our PMI results show, for example, that *narāmu* is associated with other kingly epithets in royal inscriptions and that *râmu* appears in the context of erotic love and lovemaking in the corpus of love literature. As expected, PMI succeeded in highlighting syntagmatic relationships between words, such as those between *narāmtu* and its collocates *ḫīrtu*, "(equal-ranking) wife"; *Aya* (a goddess); and *kallātu*, "daughter-in-law; bride," in royal inscriptions.

Second, we used fastText to detect words that are semantically similar to the target word, although they do not necessarily appear in the same context. A group of words that share such paradigmatic relationships consists of *râmu*, *narāmu*, *rāʾimu*, and *palāḫu*, "to fear; revere," in royal inscriptions. The words are two sides of the same coin: the three love words are used to describe the divine love of the king's correct behavior, whereas *palāḫu* expresses the king's reverence for the gods. However, the fastText results are similar to those of PMI in many cases. This

is expected in a relatively small dataset in which a cluster of highly similar contexts can dictate the semantics of a word from a quantitative perspective (see Svärd et al. 2021a). In a larger dataset, fastText would allow us to identify synonymous or similar words more accurately.

We visualized the raw results of PMI and fastText as networks by using Gephi software. This allowed us to study the relationships between several lexemes at once and observe patterns that would remain unnoticed in the raw data. Finally, we used the corpus search tool Korp to study the attestations of a single word or several co-occurring words in context. Because our dataset was small, the aggregate view given by PMI and fastText needed to be complemented by careful philological work. Korp provides easy access to the texts and allows the user to do simple and complex searches; filter the attestations according to genres, time periods, and other metadata; and study the words in their original textual context. Our research moved in a hermeneutic circle between the raw PMI and fastText results, their visualizations in Gephi, and the close reading of the texts in Korp. The data used for and created during our research is available online at https://doi.org/10.5281/zenodo.5861579, and we hope it sparks future research on this topic.

Our key findings can be summarized as follows. In royal inscriptions, *râmu*, *narāmu*, and *rāʾimu* express divine fondness for the king, who exercises his royal duties diligently, taking care of the gods and listening to their commands. At the same time, *narāmtu* and *ruʾāmu* are used to express love between divine or earthly spouses and occasionally between other family members as well. Our results in the group of literary text genres are dominated by love literature, in which *râmu*, *rāʾimu*, and *raʾīmu* denote sexual and erotic love between a man and a woman or a god and a goddess. The derivative *narāmu* is used both in love literature and in other literary text genres to express affection between spouses or other family members. In letters, *râmu* and its derivative *rāʾimu* denote instrumental love, loyalty, and obligations between mortals as well as divine love for the king. In general, our results agree with philological studies on *râmu* in these three genres, supporting the validity of our methodological approach (Bach, this volume; Podany, this volume; Wasserman 2016).

The results show that the contexts in which *râmu* and its derivatives appear are largely genre specific, although the usage of different derivatives also varies within a single genre. In the network created with fastText, emotion verbs and their derivatives are primarily clustered together according to the genre in which they are used, not according to the emotion they represent (Figure 3.8). In other words, emotion words used in royal inscriptions form a group instead of love words from different genres forming a group. Within a genre, *râmu* and its derivatives tend to be clustered together. This highlights the importance of genre for further lexical semantic work on Akkadian. Furthermore, it suggests that we may not need to rely on external labeling by modern scholars to identify genres, but we could use computational analysis of the texts themselves to identify groups which would have made sense from the perspective of a Mesopotamian scribe.

The network in Figure 3.8 was produced from the lists of emotion words and the ten words most similar to them. In the figure, only the emotion words are displayed, and they are labeled according to the genre in which they appear (2 astrological/astronomical, 3 grant/decree/gift, 4 legal transaction, 5 letter, 7 literary, 8 miscellaneous, 9 omen/divination, 10 prayer/ritual/incantation, 11 royal inscription, and 12 scholarly). The word *râmu* and its derivatives are underlined.

Our methods and workflow can be applied to any lemmatized textual dataset which is large enough to benefit from statistical methods. We use words that occur in the same context to characterize the meaning of a lexeme – in this case an emotion word. This is based on the premise that the meaning of a word depends on the other words that occur in the same context (Firth 1957; Nida 2001, 31–36), and thus the co-occurring words form a cloud of associations which explicate the emotion for the contemporary user of the language. The native speakers of Akkadian are long gone, but statistical analysis provides a promising method to study

Figure 3.8 Network showing the clustering of emotion words according to the genre in which they appear, created using cosine similarities from fastText as edge weights.

these associations in the surviving text corpus. Such statistical analysis has implications for understanding precise nuances of words, which are often lost in translation. The future aim of our research team is to make extensive linguistic networks of Akkadian words available to all scholars to facilitate and speed up their philological work.

Notes

1 We gratefully acknowledge that the research for this chapter has been funded by the Academy of Finland (decision numbers 298647, 312051, and 330727). We thank the Open Richly Annotated Cuneiform Corpus (Oracc) for their efforts in making linguistically annotated cuneiform texts available online. We are indebted to everyone who has been involved in creating this research data, including the authors of the original publications and the researchers who have made the data Oracc compatible and enriched it through lemmatizations and by adding other metadata (for a list of projects and their contributors, see the file OraccCredits.txt in our online repository at

https://doi.org/10.5281/zenodo.5861579). In the context of this chapter, we want to acknowledge the work of the Munich Open-access Cuneiform Corpus Initiative (PIs Karen Radner and Jamie Novotny), the Royal Inscriptions of the Neo-Assyrian Period project (PI Grant Frame), and the Akkadian Love Literature project (Nathan Wasserman and Yigal Bloch) in particular. We thank Johannes Bach and Amanda H. Podany for sharing their unpublished work with us and our colleagues at the Centre of Excellence in Ancient Near East Empires for their comments and feedback on this chapter. We acknowledge FIN-CLARIN and the Language Bank of Finland for hosting the data and the content search system. Finally, we are grateful to Albion M. Butters for revising the English language of the chapter. Work on this chapter was jointly conducted by all the authors, but for the most part, the division of work was as follows: Alstola wrote the sections "Introduction," "Availability of and Prerequisites for Data," "Network Analysis and Visualization," "Workflow," "Royal Inscriptions," and "Literary Text Genres"; Jauhiainen wrote "Dataset for This Chapter," "Continuous Skip-Gram Model Implemented Using FastText," and "Corpus Search Tool Korp"; Sahala wrote "Pointwise Mutual Information with Context Similarity Weighting"; Alstola, Jauhiainen, and Sahala wrote "Word Similarity"; Alstola, Lindén, and Svärd wrote "Conclusions"; Alstola and Svärd wrote "Introduction" (to the case study of *râmu*) and "Letters"; and Jauhiainen and Lindén wrote "Methods and Approaches" (the authors are listed in alphabetical order). Alstola coordinated the design for the workflow, and Alstola and Svärd analyzed the results. Jauhiainen processed the dataset. Sahala designed the weighting algorithm for PMI and wrote the tool for calculating the PMI scores. Lindén directed the language-technological work.

2 http://oracc.museum.upenn.edu/.

3 For classical Greek and Latin texts, see Berti 2019 and the resources of the Scaife Viewer (https://scaife.perseus.org/). For biblical texts, see the resources of the Eep Talstra Centre for Bible and Computer (http://etcbc.nl/; https://github.com/ETCBC) and the STEP Bible data by Tyndale House (www.stepbible.org/; https://github.com/tyndale/STEPBible-Data).

4 When finishing this chapter, we discovered that a tiny number of texts with Aramaic and Persian words are included in the dataset. Since this has no real impact on the results, we decided against editing the dataset and rerunning the analysis.

5 Role and Nadif (2011) call this measure just PPMI, although it is a variant of Daille's PMI^2. As PPMI normally refers to regular PMI with all negative scores discarded, we call this measure $PPMI^2$ instead.

6 There is also a mathematical explanation for generally low $PPMI^2$ scores: the measure is derived from Daille's PMI^2 by taking a base-2 exponent function of it. Thus, for instance, PMI^2 scores between 0 and −10 correspond to exponentially decaying $PPMI^2$ scores of $2^0 = 1$ to $2^{-10} \approx 0.001$. Because very small numbers seem to cause problems for Gephi, we take the square root of the $PPMI^2$ scores before using them as edge weights in our networks.

7 In Sahala and Lindén 2020, the method is improved and generalized by directly weighting the co-occurrence frequencies instead of the final scores. This makes it better applicable to a large variety of collocation measures.

8 For another example, see Svärd et al. 2021a, 481. The word *a* is always in either the first or the middle position, depending on the window symmetry.

9 https://korp.csc.fi/#?lang=en; www.kielipankki.fi/support/korp/.

10 They are numbered as follows: 1) administrative record, 2) astrological/astronomical, 3) grant/decree/gift, 4) legal transaction, 5) letter, 6) lexical, 7) literary, 8) miscellaneous, 9) omen/divination, 10) prayer/ritual/incantation, 11) royal inscription, 12) scholarly, 13) school, and 14) uncertain or unspecified.

11 The word transcribed as *irimmu* in Oracc, CAD, and CDA is the same word as *irʾemu* (eSAD). The meaning "love charm" is given in both eSAD and Wasserman 2016, 53. As this derivative of *râmu* appears 11 times in the Oracc corpus of love literature, it would have deserved an analysis with PMI and fastText. However, it was detected only in the very last stages of writing this chapter and could not be studied more closely.

Bibliography

Alstola, Tero, Heidi Jauhiainen, Saana Svärd, Aleksi Sahala, and Krister Lindén. 2022. Dataset used for "Digital Approaches to Analyzing and Translating Emotion: What Is Love?" *Zenodo*. https://doi.org/10.5281/zenodo.5861580.

Alstola, Tero, Shana Zaia, Aleksi Sahala, Heidi Jauhiainen, Saana Svärd, and Krister Lindén. 2019. "Aššur and His Friends: A Statistical Analysis of Neo-Assyrian Texts." *Journal of Cuneiform Studies* 71: 159–80. https://doi.org/10.1086/703859.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. 2014. "Don't Count, Predict! A Systematic Comparison of Context-Counting vs. Context-Predicting Semantic Vectors." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 1, *Long Papers*, edited by Kristina Toutanova and Hua Wei, 238–47. Stroudsburg: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-1023.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, edited by Eytan Adar, Matthew Hurst, Tim Finin, Natalie Glance, Nicolas Nicolov, and Belle Tseng, 361–62. Menlo Park: AAAI Press. https://ojs.aaai.org/index.php/ICWSM/article/view/13937.

Berti, Monica, ed. 2019. *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Age of Access? Grundfragen der Informationsgesellschaft 10. Berlin: De Gruyter Saur. https://doi.org/10.1515/9783110599572.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5: 135–46. https://doi.org/10.1162/tacl_a_00051.

Borin, Lars, Markus Forsberg, and Johan Roxendal. 2012. "Korp – The Corpus Infrastructure of Språkbanken." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 474–78. European Language Resources Association. www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf.

Canales, Lea, and Patricio Martínez-Barco. 2014. "Emotion Detection from Text: A Survey." In *Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC)*, edited by Myriam Hernandez and Josafá de Jesus Aguiar Pontes, 37–43. Stroudsburg: Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-6905.

Chandler, Daniel. 2007. *Semiotics: The Basics*. Second edition. London: Routledge.

Church, Kenneth W., and Patrick Hanks. 1990. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics* 16 (1): 22–29. www.aclweb.org/anthology/J90-1003.

Cong, Jin, and Haitao Liu. 2014. "Approaching Human Language with Complex Networks." *Physics of Life Reviews* 11: 598–618. https://doi.org/10.1016/j.plrev.2014.04.004.

Daille, Béatrice. 1994. *Approche mixte pour l'extraction automatique de terminologie: Statistiques lexicales et filtres linguistiques*. PhD disseration, Université Paris 7.

Da Riva, Rocío. 2013. *The Inscriptions of Nabopolassar, Amēl-Marduk and Neriglissar. Studies in Ancient Near Eastern Records 3*. Boston: de Gruyter. https://doi.org/10.1515/9781614513551.

Elwert, Frederik, and Simone Gerhards. 2017. "Tracing Concepts – Semantic Network Analysis as a Heuristic Device for Classification." In *Classification from Antiquity to Modern Times: Sources, Methods, and Theories from an Interdisciplinary Perspective*, edited by Tanja Pommerening and Walter Bisang, 311–38. Berlin: de Gruyter. https://doi.org/10.1515/9783110538779-012.

Fales, Frederick Mario. 2000. "bīt-bēli: An Assyrian Institutional Concept." In *Patavina Orientalia Selecta (History of the Ancient Near East/Monographs 4)*, edited by Elena Rova, 231–49. Padova: Sargon.

Firth, John R. 1957. "A Synopsis of Linguistic Theory, 1930–55." In *Studies in Linguistic Analysis*, edited by John R. Firth, 1–32. Oxford: Blackwell.

Foster, Benjamin R. 2007. *Akkadian Literature of the Late Period*. Guides to the Mesopotamian Textual Record 2. Münster: Ugarit-Verlag.

Ge, Jing, Marisol Alonso Vazquez, and Ulrike Gretzel. 2018. "Sentiment Analysis: A Review." In *Advances in Social Media for Travel, Tourism and Hospitality: New Perspectives, Practice and Cases*, edited by Marianna Sigala and Ulrike Gretzel, 243–61. New Directions in Tourism Analysis. London: Routledge.

Halton, Charles, and Saana Svärd. 2018. *Women's Writing of Ancient Mesopotamia: An Anthology of the Earliest Female Authors*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107280328.

Jackson, Joshua C., Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray, and Kristen A. Lindquist. 2019. "Emotion Semantics Show Both Cultural Variation and Universal Structure." *Science* 366: 1517–22. https://doi.org/10.1126/science.aaw8160.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9: e98679. https://doi.org/10.1371/journal.pone.0098679.

Jaques, Margaret. 2006. *Le vocabulaire des sentiments dans les textes sumériens: Recherche sur le lexique sumérien et akkadien (Alter Orient und Altes Testament 332)*. Münster: Ugarit-Verlag.

Jauhiainen, Heidi, and Tero Alstola. Forthcoming. "Fast(Text) Analysis of Mesopotamian Divine Names." In *The Ancient World Goes Digital (CyberResearch vol. 2): Case Studies on Archaeology (Landscape Archaeology and Artefacts), Texts, Online Publishing, Digital Archiving, and Preservation*, edited by Vanessa Bigot Juloux, Alessandro di Ludovico, Sveta Matskevich, and Randall W. Younker. Digital Biblical Studies. Leiden: Brill.

Jauhiainen, Heidi, Aleksi Sahala, and Tero Alstola. 2019. *Open Richly Annotated Cuneiform Corpus, Korp Version, May 2019*. Kielipankki. http://urn.fi/urn:nbn:fi:lb-2019060601.

Jauhiainen, Heidi, Aleksi Sahala, Tero Alstola, Saana Svärd, and Krister Lindén. 2021. *Lexicographical Portal – The Dataset*. Zenodo. http://doi.org/10.5281/zenodo.4646662.

Jurafsky, Daniel, and James H. Martin. 2019. *Speech and Language Processing, 3rd ed., chapter 6: Vector Semantics and Embeddings*. Unpublished Manuscript. https://web.stanford.edu/~jurafsky/slp3/.

Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics* 3: 211–25. https://transacl.org/index.php/tacl/article/view/570/124.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, edited by Christopher J. C. Burges, Léon Bottou, Max Welling, Zoubin Ghahramani, and Kilian Q. Weinberger, 3111–19. https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781. https://arxiv.org/abs/1301.3781.

Newman, Mark. 2018. *Networks*. Second edition. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198805090.001.0001.

Nida, Eugene A. 2001. *Contexts in Translating (Benjamins Translation Library 41)*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.41.

Quispe, Laura V. C., Jorge A. V. Tohalino, and Diego R. Amancio. 2021. "Using Virtual Edges to Improve the Discriminability of Co-Occurrence Text Networks." *Physica A* 562: 125344. https://doi.org/10.1016/j.physa.2020.125344.

Role, François, and Mohamed Nadif. 2011. "Handling the Impact of Low Frequency Events on Co-occurrence Based Measures of Word Similarity: A Case Study of Pointwise Mutual Information." In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR 2011)*, edited by Joaquim Filipe and Ana Fred, 226–31. https://doi.org/10.5220/0003655102260231.

Sahala, Aleksi. 2019. "Pmizer: A Tool for Calculating Word Association Measures." *Github*. https://github.com/asahala/Pmizer.

Sahala, Aleksi, and Krister Lindén. 2020. "Improving Word Association Measures in Repetitive Corpora with Context Similarity Weighting." In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020)*, vol. 1, *KDIR*, edited by Ana Fred and Joaquim Filipe, 48–58. https://doi.org/10.5220/0010106800480058.

Schaudig, Hanspeter. 2001. *Die Inschriften Nabonids von Babylon und Kyros' des Großen samt den in ihrem Umfeld entstandenen Tendenzschriften: Textausgabe und Grammatik (Alter Orient und Altes Testament 256)*. Münster: Ugarit-Verlag.

Schnoebelen, Tyler J. 2012. *Emotions Are Relational: Positioning and the Use of Affective Linguistic Resources*. PhD Disseration, Stanford University. http://purl.stanford.edu/fm335ct1355.

Svärd, Saana, Tero Alstola, Heidi Jauhiainen, Aleksi Sahala, and Krister Lindén. 2021a. "Fear in Akkadian Texts: New Digital Perspectives on Lexical Semantics." In *The Expression of Emotions in Ancient Egypt and Mesopotamia (Culture and History of the Ancient Near East 116)*, edited by Shih-Wei Hsu and Jaume Llop Raduà, 470–502. Leiden: Brill. https://doi.org/10.1163/9789004430761_019.

Svärd, Saana, Tero Alstola, Heidi Jauhiainen, Aleksi Sahala, and Krister Lindén. 2021b. Dataset used for "Fear in Akkadian Texts: New Digital Perspectives on Lexical Semantics." Zenodo. https://doi.org/10.5281/zenodo.3634324.

Svärd, Saana, Heidi Jauhiainen, Aleksi Sahala, and Krister Lindén. 2018. "Semantic Domains in Akkadian Texts." In *CyberResearch on the Ancient Near East and Neighboring Regions: Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving (Digital Biblical Studies 2)*, edited by Vanessa Bigot Juloux, Amy R. Gansell, and Alessandro Di Ludovico, 224–56. Leiden: Brill. https://doi.org/10.1163/9789004375086_009.

Toivonen, Riitta, Mikko Kivelä, Jari Saramäki, Mikko Viinikainen, Maija Vanhatalo, and Mikko Sams. 2012. "Networks of Emotion Concepts." *PLoS ONE* 7(1):e28883. https://doi.org/10.1371/journal.pone.0028883.

Wasserman, Nathan. 2016. *Akkadian Love Literature of the Third and Second Millennium BCE (Leipziger Altorientalistische Studien 4)*. Wiesbaden: Harrassowitz. https://doi.org/10.2307/j.ctvckq4rk.

Watanabe, Chikako E. 2002. *Animal Symbolism in Mesopotamia: A Contextual Approach (Wiener Offene Orientalistik 1)*. Vienna: Institut für Orientalistik der Universität Wien.